

Exploratory Item Classification Via Spectral Graph Clustering

Applied Psychological Measurement

2017, Vol. 41(8) 579–599

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0146621617692977

journals.sagepub.com/home/apm



Yunxiao Chen¹, Xiaou Li², Jingchen Liu³,
Gongjun Xu⁴, and Zhiliang Ying³

Abstract

Large-scale assessments are supported by a large item pool. An important task in test development is to assign items into scales that measure different characteristics of individuals, and a popular approach is cluster analysis of items. Classical methods in cluster analysis, such as the hierarchical clustering, K-means method, and latent-class analysis, often induce a high computational overhead and have difficulty handling missing data, especially in the presence of high-dimensional responses. In this article, the authors propose a spectral clustering algorithm for exploratory item cluster analysis. The method is computationally efficient, effective for data with missing or incomplete responses, easy to implement, and often outperforms traditional clustering algorithms in the context of high dimensionality. The spectral clustering algorithm is based on graph theory, a branch of mathematics that studies the properties of graphs. The algorithm first constructs a graph of items, characterizing the similarity structure among items. It then extracts item clusters based on the graphical structure, grouping similar items together. The proposed method is evaluated through simulations and an application to the revised Eysenck Personality Questionnaire.

Keywords

spectral clustering, cluster analysis, large-scale assessment, personality assessment, Eysenck Personality Questionnaire

Introduction

With the rapid developments in information technology during the past several decades, Internet-based and mobile-enabled psychological assessments are multiplying. As a result, large item pools tend to emerge, together with availability of a large amount of response data. An

¹Emory University, Atlanta, GA, USA

²University of Minnesota, Minneapolis, USA

³Columbia University, New York, NY, USA

⁴University of Michigan, Ann Arbor, USA

Corresponding Author:

Yunxiao Chen, Department of Psychology, Emory University, 36 Eagle Row #270, Atlanta, GA 30322, USA.

Email: yunxiao.chen@emory.edu

important question then becomes how to make use of such rich resource to provide better psychological measurement. As a first step to answering this question, the authors propose a data-driven approach to assign a large number of items into clusters in such a way that items within each cluster are homogeneous, while items in different clusters are heterogeneous. One application of the proposed method is in personality assessments, a psychological assessment that is widely used to assess aspects of an individual's character, such as in relationship counseling, career counseling, and employment testing. Due to the heterogeneity of personality, its assessments traditionally have large item pools, typically on the scale of hundreds. In recent years, even larger item pools become available, and large-scale response data are collected. For example, the International Personality Item Pool (IPIP; Goldberg et al., 2006) contains more than 3,000 items. The proposed method analyzes such large item pools by decomposing them into scales of items that are much easier to understand and interpret.

In this study, the authors propose a method to answer two questions: How many clusters should be formed among a set of items and how to assign each item to a cluster? In particular, they propose a classification method for item analysis based on a spectral clustering algorithm. The proposed method is capable of handling high-dimensional responses and admits a low computational overhead; for instance, a maximum of 500 items is considered in the simulation study in Online Appendix A. Unlike many traditional clustering methods, this one can be effectively used when data have missing or incomplete responses that is often the case when analyzing a large item pool. The response type can be either continuous or ordinal. And, consequently, the method can cover a wide range of situations. Moreover, the authors propose methods to systematically choose various tuning parameters, among which the most important one is the number of clusters. The proposed method assumes no prior information of the item structure and thus is exploratory.

In psychological assessment, people typically use latent variable models for analyzing item response data. The authors emphasize that the proposed method is a useful complement to, not a replacement of, the latent factor models. One common practice of assigning items into scales is to analyze the loading structure from exploratory factor analysis, which usually works well. However, in exploratory factor analysis, one needs to decide how many factors to keep, which rotation to use, what threshold to truncate the small loadings, and so on, which typically requires an experienced data analyst and can be labor intensive when the number of items is large. However, little human intervention is needed for the proposed method and thus can be used as a first step in data analysis. To further use an item cluster as an assessment scale, the items should be confirmed by domain experts and further analyzed by latent factor models. One may fit unidimensional latent factor models to each item cluster and conduct standard analysis, such as goodness of fit and dimensionality check, so that the measurement can still be backed by a latent factor model.

Clustering/classification is a common technique for exploratory data analysis used in psychology, image analysis, bioinformatics, and many others (Blashfield, 1980; Blashfield & Aldenderfer, 1988; Hartigan, 1975; Jain, Murty, & Flynn, 1999; Xing et al., 2010). In psychological and educational measurement, cluster analysis has been used to classify individuals into groups (Chiu & Douglas, 2013; Chiu, Douglas, & Li, 2009; Park & Lee, 2011; Willse, Henson, & Templin, 2007). Cluster analysis has also been widely used for exploratory item analysis as an alternative to factor analysis (e.g., Borgen & Barnett, 1987; Roussos, Stout, & Marden, 1998; Zhang, 2007, 2013; Zhang & Stout, 1999). In particular, learning the loading structure in item response theory models and diagnostic classification models (Chen, Liu, Xu, & Ying, 2015; J. Liu, Xu, & Ying, 2012, 2013; Sun, Chen, Liu, Ying, & Xin, 2016) is a special type of item cluster analysis that is based on probabilistic models. As pointed out by Loevinger, Gleser, and DuBois (1953), unlike factor analysis, which weights items by their factor loadings, cluster

analysis solves the practical problem of assigning items to scales on an all or none basis. In the current context, item clustering is to assign J items into M groups performed on an N -dimensional feature space, where N is the total number of respondents or the sample size. When both J and N are large, substantial computational overhead occurs for many clustering methods (Jain et al., 1999). In addition, probability-model-based methods (e.g., latent-class models) often suffer from model lack of fit for high-dimensional features. For such models, the local independence is often assumed, that is, the features are independent given the latent-class membership. Such an assumption is often violated in the presence of high-dimensional data (Chen, Li, Liu, & Ying, 2016). Moreover, the usage of most traditional clustering methods requires complete data. Upon having missing data, mainstream approaches reduce the problem to a complete data formulation by deleting missing cases or imputation, for which the costs are often significant (Chi, Chi, & Baraniuk, 2016).

Spectrum-based item clustering methods foot on graph theory, a branch of mathematics that studies the properties of graphs. A spectral clustering algorithm first constructs a graph of items based on a similarity measure defined on each pair of items, where items more similar to each other tend to be connected. Then, by spectral analysis of the graph, it produces clusters so that items within a cluster are densely connected and different clusters are loosely connected. Spectral methods admit both sound theoretical properties under regularity conditions and reliable empirical performance (Chung, 1997; Ng, Jordan, Weiss, 2002; Rohe, Chatterjee, & Yu, 2011; Shi & Malik, 2000). Such methods have been successfully applied to many fields, including image segmentation (Shi & Malik, 2000), text mining (Dhillon, 2001), and network analysis (Rohe et al., 2011).

In most applications mentioned above, the features take continuous values, and similarity measures are constructed naturally via the Euclidean distance in R^N . In the context of assessment/measurement, the responses are often binary or ordinal, for which Euclidean distance is inappropriate. In this article, the authors propose a similarity matrix based on a nonparametric statistic known as Kendall's tau (Kendall & Gibbons, 1990), which is applicable to all ordinal variables that may take different numbers of values. The calculation of the Kendall's tau statistic for item pairs can be done in parallel, further adding to its computational efficiency.

In the analysis, an important task is to assess the number of clusters. In the literature of spectral clustering, eigengap heuristic is typically used that is based on the distribution of the eigenvalues of the Laplacian matrix (von Luxburg, 2007). However, a "gap" of the eigenvalue distribution is not always straightforward to identify. The authors propose a method to select the number of clusters based on the eigenvalue distribution, motivated by parallel studies in factor analysis (Horn, 1965) and the gap statistic (Tibshirani, Walther, & Hastie, 2001). In particular, they consider the distribution of the eigenvalues, when there is no cluster structure, as the *reference distribution*, that can be obtained by Monte Carlo method. The idea is to find the eigengap by comparing the observed eigenvalues with a reference distribution obtained by simulation. This method performs well in the authors' simulation studies.

The authors apply the method to the revised Eysenck Personality Questionnaire (S. B. Eysenck, Eysenck, & Barrett, 1985), one of the most popular personality assessments. This questionnaire is derived according to Eysenck's theory on personality, which contains three scales measuring three aspects of personality: Extraversion/Introversion, Neuroticism/Stability, and Psychoticism/Socialization. The authors' analysis validates the three scales and explores finer test structure driven by data.

The rest of the article is organized as follows: The spectral clustering algorithm and the method for choosing the number of clusters are presented in the section "Spectral Clustering," and are illustrated via simulated examples in the section "Simulated Examples." The proposed method is further investigated by a real data example in the section "Application to EPQ-R

Data.” The section “Summary and Discussion” provides a summary and further discussions. Simulation studies, theory on spectral clustering, and a review of the K-means algorithm are presented in the online appendix.

Spectral Clustering

Spectral Clustering Algorithm for Items

Consider N individuals, each of whom responds to J items. Let R_{ij} be the response of individual i to item j and $R = (R_{ij})_{N \times J}$ be the entire data matrix. The responses are ordinal. Without loss of generality, assume $R_{ij} \in \{1, 2, \dots, n_j\}$. The responses to item j are further denoted by $R_j = (R_{1j}, \dots, R_{Nj})$.

A similarity matrix $S = (S_{jj'})_{J \times J}$ is introduced that is a symmetric matrix of nonnegative entries constructed based on the data, that is, $S_{jj'} = S_{j'j} \geq 0$ for all j and j' . The construction of S will be discussed in the section “Construction of the Similarity Matrix.” The spectral clustering algorithm aims to classify items into clusters, so that items in the same cluster tend to be similar (i.e., $S_{jj'}$ tends to take a large value if items j and j' belong to the same cluster) and those in different clusters tend to be dissimilar to each other (i.e., $S_{jj'}$ tends to be small if j and j' belong to different clusters). In this article, the authors focus on similarity matrices whose entries are binary ($S_{jj'} \in \{0, 1\}$), in which case S can be viewed as the adjacency matrix of an undirected graph. That is, each vertex in the graph corresponds to an item, and there is an edge between items j and j' if and only if $S_{jj'} = 1$.

A description of the spectral clustering algorithm (Shi & Malik, 2000) is provided, followed by explanations. First, suppose that the similarity graph S is given and the number of clusters M is known. The construction of S and the choices of M are discussed later.

Algorithm 1 The similarity matrix S and the number of clusters M are given.

1. Construct a diagonal matrix D and a Laplacian matrix L as follows:

$$D_{jj} = \sum_{j'=1}^J S_{jj'} \quad \text{and} \quad L = D^{-1}S.$$

2. Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_J$ be the eigenvalues of L and X_1, X_2, \dots, X_J be the corresponding eigenvectors. Let

$$X = (X_1, \dots, X_M)_{J \times M}$$

be the $J \times M$ matrix containing the first M eigenvectors.

3. The matrix X is written in rows, and let $Y_1, \dots, Y_J \in \mathbb{R}^M$ be the row vectors of X , that is

$$X = (Y_1, Y_2, \dots, Y_J)^\top.$$

Consider Y_1, \dots, Y_J as the data and apply K-means algorithm (Algorithm 1 in the online appendix) to Y_1, \dots, Y_J , assuming that there are M clusters. Let $\hat{\pi}_j \in \{1, \dots, M\}$ be the classification of Y_j .

4. Output: $\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_J$, that is, if $\hat{\pi}_j = m$, then item j is classified to cluster m .

In what follows, the authors describe heuristically the underlying mechanism of the spectral clustering algorithm. They consider an example of 10 items falling into three clusters, the graphical representation of which is shown in Figure 1. These items fall into three clusters naturally, that is, $\{1, 2, 3\}$, $\{4, 5, 6, 7\}$, and $\{8, 9, 10\}$, where items within a cluster are connected via a path and items from different clusters are disconnected. In this case, the similarity matrix S is block diagonal, as shown in the heat map in the right panel of Figure 1, where the values 0 and

1 are indicated in white and blue, respectively. For this similarity matrix, the X matrix obtained in Step 2 of the algorithm is as follows:

$$X = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}^T. \quad (1)$$

Let Y_1, \dots, Y_{10} be the row vectors of X . In particular, Y_1, Y_2 , and Y_3 are $(1/\sqrt{3}, 0, 0)^T$, Y_4 through Y_7 are $(0, 1/\sqrt{4}, 0)^T$, and Y_8, Y_9, Y_{10} are $(0, 0, 1/\sqrt{3})^T$. Thus, the K-means algorithm, which minimizes the within-cluster sum of squares, assigns items j and j' to the same cluster if $Y_j = Y_{j'}$. The spectral clustering algorithm is developed based on the above observation.

The above heuristic argument can be made general through results in graph theory. To do so, some concepts need to be introduced. Consider an undirected graph $G=(V, E)$, where $V=\{1, \dots, J\}$ is the set of vertices and $E \subset V \times V$ is the set of edges. The graph is undirected in the sense that $(i, j) \in E$ implies $(j, i) \in E$. Furthermore, let S be the adjacency matrix of G . A subset $A \subset V$ is said to be *connected* if any two vertices in A can reach each other through a path belonging to the set A . A subset A is called a *connected component* if it is connected, and there are no connections between A and A^c , such as subsets $\{1, 2, 3\}$, $\{4, 5, 6, 7\}$, and $\{8, 9, 10\}$ in the left panel of Figure 1. Each connected component is viewed as a cluster, as connectivity indicates similarity. For a graph with M connected components, if we input its adjacency matrix S and M into Algorithm 1, spectral graph theory (in particular, Theorem 1 in Online Appendix B) guarantees that $Y_i = Y_j$ if i and j are from the same connected component and $Y_i \neq Y_j$ otherwise. As a result, the K-means method is able to recover the clusters. As Y_1, \dots, Y_J can be viewed as item-specific features characterizing their cluster memberships, they are called *graph features* of items, and the matrix X is called the *graph feature matrix*.

In practice, such an ideal, that is, perfectly separated, case, is rarely observed as the similarity matrix S is constructed with noise and the underlying cluster structure may not be perfect. However, if the observed similarity graph approximates the ideal situation, that is, items within a cluster are densely connected and different clusters are loosely connected, the spectral clustering method is able to recover the clusters, backed by matrix perturbation theory (Stewart & Sun, 1990). Readers are referred to von Luxburg (2007) for further details. This point is confirmed by simulation studies in the section “Simulated Examples” and Online Appendix A.

Construction of the Similarity Matrix

The authors now propose a method for constructing a similarity matrix S for the spectral clustering algorithm from item response data. To begin with, Kendall's tau statistic is considered as a similarity measure between two items (Howell, 2012). Let $(R_{1j}, R_{1j'}), \dots, (R_{Nj}, R_{Nj'})$ be the responses of N individuals to items j and j' . Then, Kendall's tau for item pair (j, j') is

$$\hat{\tau}_{jj'} = \frac{2}{N(N-1)} \sum_{1 \leq i < i' \leq N} \text{sgn}\{(R_{ij} - R_{i'j})(R_{ij'} - R_{i'j'})\},$$

where the function $\text{sgn}(x)$ takes value 1 if $x > 0$, 0 if $x = 0$, and -1 otherwise. Clearly, $\hat{\tau}_{jj'}$ lies between -1 and 1. Kendall's tau, which has been widely used in nonparametric statistics (Sprent & Smeeton, 2016), measures the strength of dependence. A substantial deviation from 0 means a strong correlation between the responses, and hence a high similarity between the two items. Notice that if two items are measuring a single construct but are worded reversely,

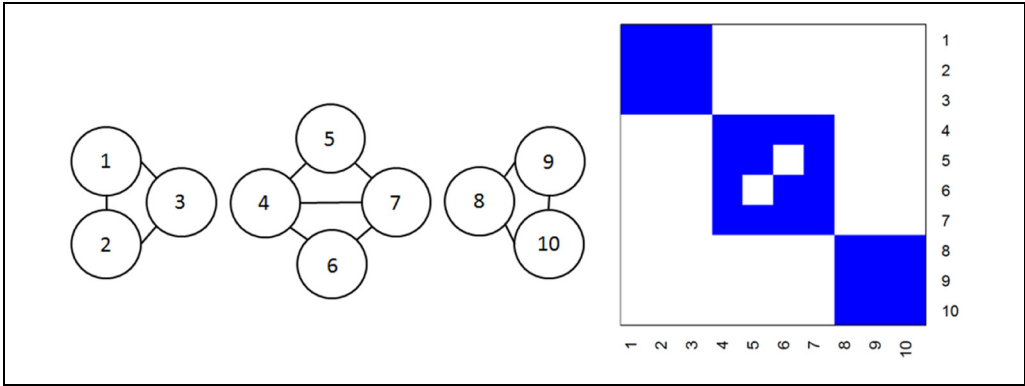


Figure 1. The graphical representation of an ideal example that each cluster is a connected component and different clusters are disconnected.

then it is likely to lead to a negative $\hat{\tau}_{jj'}$. Therefore, items j and j' are considered highly similar, if $|\hat{\tau}_{jj'}|$ takes a large value. The authors remark that $\hat{\tau}_{jj'}$ can be estimated well even when data contain missing or incomplete responses, provided that there are enough individuals responding to both items. See the section “A Brief Summary and Some Remarks” for further discussion.

To better capture the local neighborhood relations between items, a K -nearest neighbor graph is considered (von Luxburg, 2007). The set of K -nearest neighbors of an item j is first defined. Let $\{|\hat{\tau}_{jj'}|\}$ be the ordered statistics of $\{|\hat{\tau}_{jj'}| : j' \neq j\}$ in a descending order. Then, the K -nearest neighbor of item j is

$$\mathcal{N}_j(K) = \{j' : |\hat{\tau}_{jj'}| \geq |\hat{\tau}_{j(K)}|\}.$$

The similarity matrix S is defined such that j and j' are connected if j is within the K -nearest neighbors of j' or item j' is one of the K -nearest neighbors of j , that is, $S_{jj'} = 1$ if $j \in \mathcal{N}_{j'}(K)$ or $j' \in \mathcal{N}_j(K)$ and $S_{jj'} = 0$ otherwise. With this construction, S is a symmetric adjacency matrix.

Algorithm 2 Construction of similarity matrix S .

Input: Responses $(R_{ij})_{N \times J}$ and connectivity parameter K .

1. Compute the absolute values of Kendall's tau statistics $|\hat{\tau}_{jj'}|$, $j = 1, \dots, J - 1$, $j' = j + 1, \dots, J$.
2. Construct the set of K -nearest neighbor $\mathcal{N}_j(K)$ for each item j .
3. Set $S_{jj'} = 1$ if $j \in \mathcal{N}_{j'}(K)$ or $j' \in \mathcal{N}_j(K)$ and $S_{jj'} = 0$ otherwise, $j \neq j' \in \{1, 2, \dots, J\}$.

Output: Similarity matrix S .

The authors discuss the choice of connectivity parameter K . Roughly speaking, K should be chosen such that the resulting graph S is neither too sparse nor too dense. When the graph is too sparse (K being too small), S typically has many connected components, so that the clustering result may not be stable. When the graph is too dense (K being too large), clusters tend to merge, and thus the classification becomes less accurate and the number of clusters cannot be well estimated. This will be further illustrated in the simulation studies. For large graphs, as suggested by the asymptotic connectivity results (Brito, Chavez, Quiroz, & Yukich, 1997), the resulting graph is almost certainly connected when K is chosen in the order of $\log(J)$. Following the

above discussion, the authors suggest a rule of thumb for choosing K . It performs reasonably well based on results from simulation studies and real data analysis.

Algorithm 3 Choice of connectivity parameter K .

Input: Responses $(R_{ij})_{N \times J}$.

Initialize: $K = \lceil \log(J) \rceil$, where $\lceil \log(J) \rceil$ is the smallest integer greater than or equal to $\log(J)$.

1. Obtain a similarity matrix S using Algorithm 2.
2. If S constructed in Step 1 corresponds to a connected graph, stop. Otherwise, $K \leftarrow K + 1$, and repeat Steps 1 and 2.

Output: Connectivity parameter K .

Finding the Number of Clusters

A commonly used method for choosing M in the spectral clustering is the so-called eigengap heuristic (Azran & Ghahramani, 2006; von Luxburg, 2007), a method catered to the spectral clustering. It chooses the number m , so that all eigenvalues $\lambda_1, \dots, \lambda_m$ are very large, but λ_{m+1} is relatively small. This method comes from the ideal situation of M completely disconnected clusters, where the eigenvalue 1 has multiplicity M , and then there is a gap to the $(M+1)$ th eigenvalue $\lambda_{M+1} < 1$. However, in practice, a “gap” of the eigenvalue distribution is not always straightforward to identify, as shown in the simulated examples in the section “Simulated Examples.” The authors propose a method that improves on the eigengap heuristic. The method is motivated by the gap statistic (Tibshirani et al., 2001) and the parallel analysis in factor analysis (Horn, 1965).

The basic idea behind the authors’ approach is to compare the observed eigenvalues with a reference distribution, that is, the distribution of the eigenvalues when there is no cluster structure. If there are M clusters, then the top M eigenvalues of the Laplacian matrix are significantly larger than this reference distribution. Specifically, the following reference distribution is considered. Given an observed similarity matrix S with binary entries, let

$$p_S = \frac{2 \sum_{j < j'} S_{jj'}}{N(N-1)}$$

be the graph sparsity. A random similarity matrix $\tilde{S} = (\tilde{S}_{jj'})$ is considered, such that $\tilde{S}_{jj'}$ are independent and identically distributed Bernoulli random variables with success probability p_S . Let \tilde{L} be the Laplacian matrix of \tilde{S} , and $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 > \dots > \tilde{\lambda}_J$ be the eigenvalues of \tilde{L} . The joint distribution of $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_J)$ is considered as the reference distribution, denoted by F_S . For each $1 \leq j \leq J$, let \tilde{t}_j be the $1 - \alpha$ quantile of the marginal distribution of $\tilde{\lambda}_j$. The number of clusters is then chosen as

$$M = \min\{m : \tilde{t}_{m+1} \neq 1, \lambda_{m+1} < \tilde{t}_{m+1}\},$$

where λ_m is the m -th largest eigenvalue of the observed Laplacian matrix L . In their implementation, the authors choose $\alpha = 5\%$. In practice, the distribution F_S is not analytically tractable, and they obtain it by Monte Carlo. The following description summarizes the method:

Algorithm 4 Parallel analysis for choosing the number of clusters.

Input: Symmetric similarity matrix S .

1. Compute the eigenvalues $\lambda_1, \dots, \lambda_J$ of L and the sparsity of S , P_S .

2. Simulate B independent samples from the reference distribution F_S , denoted by $(\tilde{\lambda}_1^{(1)}, \dots, \tilde{\lambda}_J^{(1)}), (\tilde{\lambda}_1^{(2)}, \dots, \tilde{\lambda}_J^{(2)}), \dots, (\tilde{\lambda}_1^{(B)}, \dots, \tilde{\lambda}_J^{(B)})$.
3. Compute the 95% quantiles of the marginal distribution of $(\tilde{t}_1, \dots, \tilde{t}_J)$.
4. The minimum m is chosen, such that $\tilde{t}_{m+1} \neq 1$ and $\lambda_{m+1} < \tilde{t}_{m+1}$.

Output: The chosen number of clusters $\hat{M} = m$.

A Brief Summary and Some Remarks

The authors provide a brief summary of the proposed method as follows:

1. Choose the connectivity parameter K using Algorithm 3.
2. Given K from Step 1, construct the similarity matrix S using Algorithm 2.
3. Given S as the input, run Algorithm 4 and output the number of clusters \hat{M} .
4. Run Algorithm 1 with \hat{M} clusters and output the cluster membership $\hat{\pi}_1, \hat{\pi}_2, \dots$, and $\hat{\pi}_J$.

For advantages and special features of the proposed method, the authors have the following remarks:

1. Computation. The most computationally intensive step in Algorithm 1, the K-means step, has $O(JM^2)$ complexity per iteration, while that of directly applying the K-means clustering is $O(JMN)$. The difference is significant for a large sample size N . In addition, although spectral clustering has an additional step of constructing a similarity matrix, it can be efficiently obtained by parallel computing, that is, computing $|\tau_{jj'}|$ and $s_{jj'}$ can be paralleled in item pairs, and calculating $\mathcal{N}_j(K)$ can be paralleled in items.
2. Comparison. One traditional approach to assigning items to scales is exploratory item factor analysis, which often produces reasonable results. It typically requires human intervention, which can be labor intensive when the number of items is large. However, no human intervention is needed for the proposed method. In addition, as a nonparametric approach, the proposed method performs quite robustly under different settings in terms of choosing the number of clusters and classification, whereas exploratory factor analysis may fail due to model misspecification. See the section "Comparing With Exploratory Factor Analysis" for examples.
3. Missing data. Item response data often contain missing values. When the number of items is large, it is likely that many or even all individuals have missing responses. Most traditional clustering methods suffer from such a situation. This is because they usually throw away incomplete cases, resulting in severe information loss. Another strategy for handling missing data is imputation, which fills in missing entries with plausible estimates of their values. Imputation methods can be very complicated, because information about the joint distribution of the missingness patterns and the data is needed for effective imputation. It is nontrivial to find a reasonable imputation model, and even with a given model, the imputation process is often computationally expensive (Chi et al., 2016).

The proposed method makes better use of data in a computationally efficient manner. Let Z_{ij} indicate whether the response from individual i to item j is missing ($Z_{ij} = 0$) or not ($Z_{ij} = 1$). Then for an item pair (j, j') , Kendall's tau statistic adapted to missing data becomes

$$\tilde{\tau}_{jj'} = \frac{2}{N(N-1)} \sum_{1 \leq i < i' \leq N} Z_{ij} Z_{ij'} \text{sgn}\{(R_{ij} - R_{i'j})(R_{ij'} - R_{i'j'})\},$$

where $\tilde{N} = \sum_{i=1}^N Z_{ij} Z_{ij'}$ can be viewed as the effective sample size for item pair (j, j') . For the traditional methods, the sample size of complete cases is $\sum_{i=1}^N \prod_{j=1}^J Z_{ij}$, which is usually much smaller than \tilde{N} . For a reasonably large \tilde{N} , $\tilde{\tau}_{jj'}$ is a good estimator of its population version. In particular, it is consistent when data are missing at random. Therefore, to develop a test, it can be designed to have each individual responding to a subset of items of an item pool, making data collection easier and reduce the time cost for each respondent. It is worth noting that one should be careful when the data missingness is nonignorable.

4. Data driven. Personality scale construction starts with a large item pool, with items from multiple sources, such as textbooks in psychiatry, earlier published scales of personal and social attitude, and so on. In the early stage of inventory construction, little is known for a large number of items. Two items may measure a single construct but are reversely worded, such as items “Do you prefer to go your own way rather than act by the rules?” and “Should people always respect the law?” from the P scale of Eysenck Personality Questionnaire–Revised (EPQ-R). Similarity measure based on the Euclidian distance regards two items as dissimilar, based on which the two items are unlikely to be clustered together. The proposed method solves this issue by taking the absolute value of Kendall’ tau statistic as the similarity measure. This point will be further illustrated in the real data analysis discussed in the section “Application to EPQ-R Data.”
5. Other similarity measures. Kendall’s tau statistic is a standard measure of the association between ordinal variables, and is easy to compute. However, thanks to the flexibility of the authors’ method, other reasonable similarity measures can also be used. For example, if it is believed that the responses are obtained by truncating Gaussian random variables, the similarity may be better measured by polychoric correlation (Lee, Poon, & Bentler, 1995).
6. Relationship between clusters. When item clusters are obtained from exploratory factor analysis, each cluster is associated with a latent factor and one may quantify the similarity between the clusters by the corresponding correlation between factors. There is an analog under the proposed approach, which makes use of the constructed similarity graph of items. That is, the similarity between two item clusters could be quantified by connectivity measures, such as the number of edges between two clusters or its normalized version (Kolaczyk, 2009).

Simulated Examples

Illustration

The authors first illustrate the use of the proposed method using simulated examples from a latent factor model, which is one of the most popular models for personality test (Cattell & Kline, 1977; H. J. Eysenck, 1982). A simple structure is often believed to exist in these factor models (Cattell & Kline, 1977), that is, each item loads on only a few factors. In this section, the authors investigate the spectral clustering algorithm using two data sets generated from a latent factor model. The data admit a simple structure that each item only loads on one factor, so that the item clusters are well defined (by the latent factor that each item measures). In particular, there are M latent factors. For each individual i , let $\theta_i = (\theta_{i1}, \dots, \theta_{iM})$ be the M -

dimensional factor score. For each item j , let $\pi_j \in \{1, \dots, M\}$ be the factor it measures and a_j be the corresponding factor loading. More precisely,

$$P(R_{ij} = 1 | \theta_i) = \frac{e^{a_j \theta_i \pi_j}}{1 + e^{a_j \theta_i \pi_j}}.$$

Furthermore, each item loads on factor m with probability $P(\pi_j = m) = p_m$ for $m = 1, \dots, M$.

A setting in which there are $M = 3$ correlated latent factors is considered. The factor score for each individual is

$$\theta_i \sim N(0, \Sigma),$$

where $\Sigma_{mm} = 1, m = 1, \dots, M$ and $\Sigma_{mm'} = 0.2, m \neq m'$. The factor loading a_j s are independently sampled from uniform distribution over interval $[0.1, 2]$. The number of items $J = 60$ and $p_m = \frac{1}{M}, m = 1, \dots, M$. In addition, two sample sizes are considered: $n = 100$ and $n = 400$.

The authors construct the similarity matrix corresponding to $K = 10$ nearest neighbors. The similarity matrices are visualized via heat maps in Figure 2, in which blue cells indicate an edge between two items. They also circle items in the same cluster. There are more within-class edges than between-class edges. In addition, the top eigenvalues are displayed in Figure 3. For the case $n = 400$, the largest three eigenvalues are close to 1.

The authors illustrate the performance of the spectral clustering algorithm for these two data sets. The vectors Y_1, \dots, Y_J corresponding to the second and the third eigenvectors are displayed in Figure 4. The authors did not display the first eigenvector because it is a vector of all ones and contains no information (Proposition 1 in Online Appendix B). In the figure, the true class memberships are labeled by different colors. Cluster boundaries identified by the K-means algorithm are the black lines. For both data sets, the Y vectors admit the same cluster structure as that of the items. The proposed method performs well, even if the similarity matrices contain certain level of noise. There are six items misclassified for the case $n = 100$ and one item misclassified for the case $n = 400$.

To illustrate the sensitivity to the choice of the connectivity parameter K , the authors consider the data set with $n = 100$ and various connectivity parameter values $K = 5, 10, 20$, and 30 . Note that $K = 5$ is the choice based on Algorithm 3. The corresponding similarity matrices and the results of clustering are shown in Figures 5 and 6. According to Figure 5, as K increases, the number of edges increases. Nonetheless, there are still more within-cluster edges than between-cluster edges. In addition, according to Figure 6, the Y vectors corresponding to different clusters are reasonably separated. Therefore, the spectral clustering algorithm is reasonably robust to the choice of K , even for a large connectivity parameter value that leads to a dense graph (e.g., when $K = 30$). The numbers of misclassified items are small, that is, 7, 6, 5, and 5, respectively, for $K = 5, 10, 20$, and 30 . The sensitivity to the choice of K on clustering accuracy is further studied in the simulation studies in Online Appendix A.

The authors further illustrate the proposed parallel analysis for selecting M . They consider several connectivity parameters $K = 5, 10, 20$, and 30 . Figures 7 and 8 show the observed eigenvalues and the corresponding thresholds computed based on the Monte Carlo samples from the reference distribution. All these plots start from the second largest eigenvalue, because the largest eigenvalue is always 1 (Proposition 1 in Online Appendix B). According to the results, the proposed method correctly selects the number of factors under all K . The influence of K on the parallel analysis will be further evaluated in the simulation studies and real data analysis.

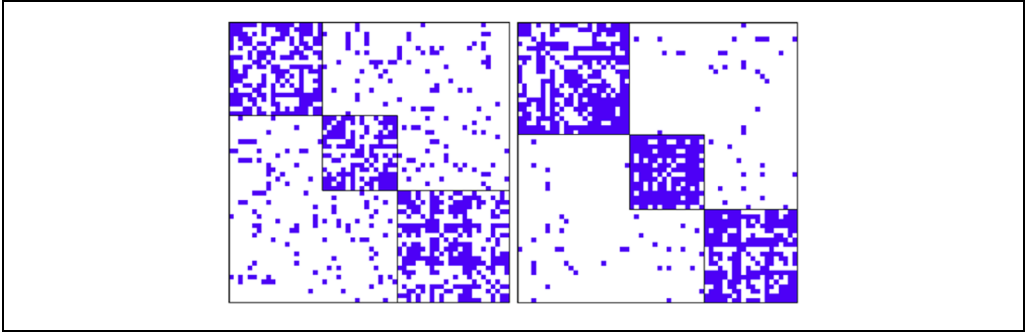


Figure 2. Similarity matrix S for $K = 10$ for sample size $n = 100$ (left) and $n = 400$ (right).

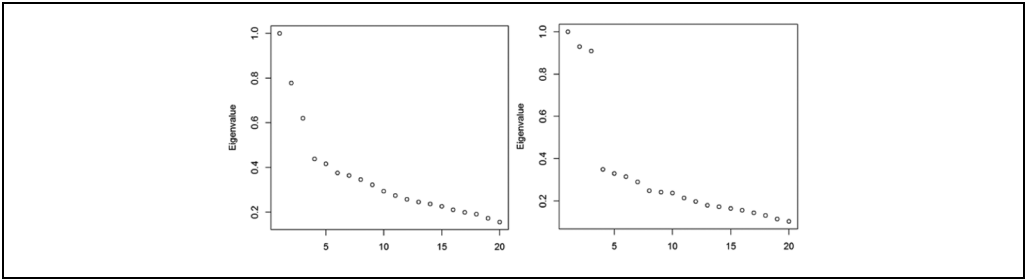


Figure 3. Top 20 eigenvalues of the Laplacian matrix L .
Note. Left panel: $n = 100$; right panel: $n = 400$.

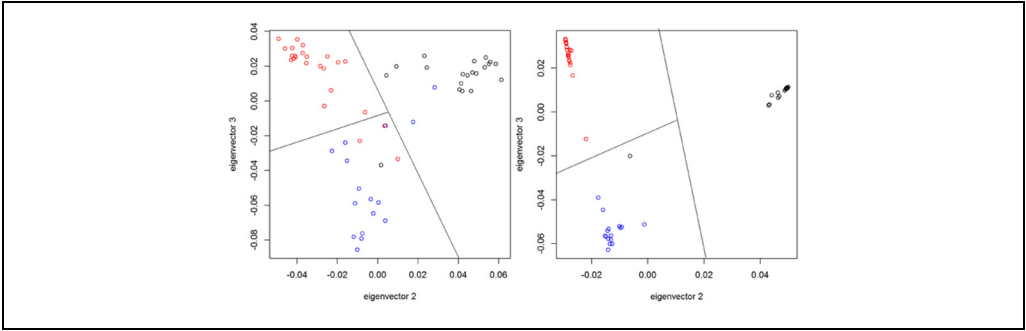


Figure 4. Second and third eigenvectors.
Note. Left panel: $n = 100$; right panel: $n = 400$.

For these two data sets, if K-means algorithm is directly applied (given the number of clusters), the numbers of misclassified items are 5 and 1, respectively, which is similar to the result of spectral clustering. However, if there exist reversely worded items, the K-means algorithm performs poorly. For example, the authors randomly choose half of the items and flip the responses, so that these items can be viewed as reversely worded items. The number of misclassified items for the two data sets then increases to 30 and 22, respectively. This is because the K-means algorithm measures the similarity between data points by their Euclidian distance.

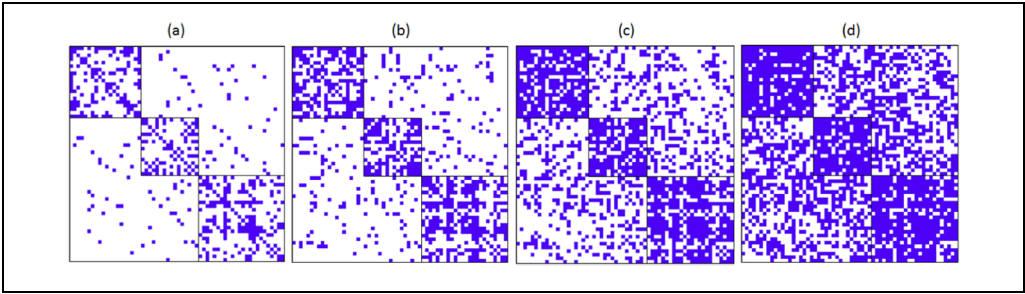


Figure 5. Constructed similarity matrix S for different K .
Note. (a) $K = 5$; (b) $K = 10$; (c) $K = 20$; (d) $K = 30$.

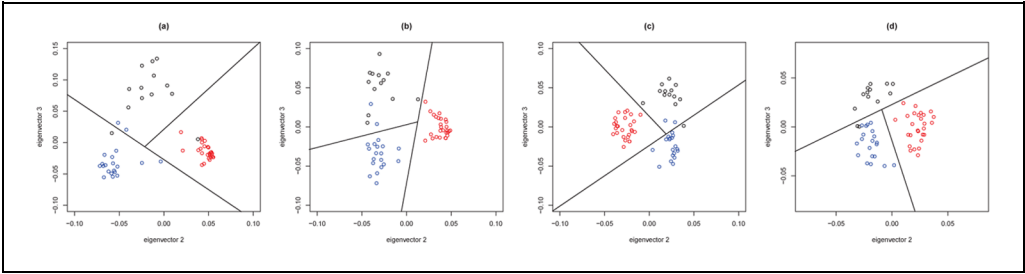


Figure 6. Result of clustering for different K .
Note. (a) $K = 5$; (b) $K = 10$; (c) $K = 20$; (d) $K = 30$.

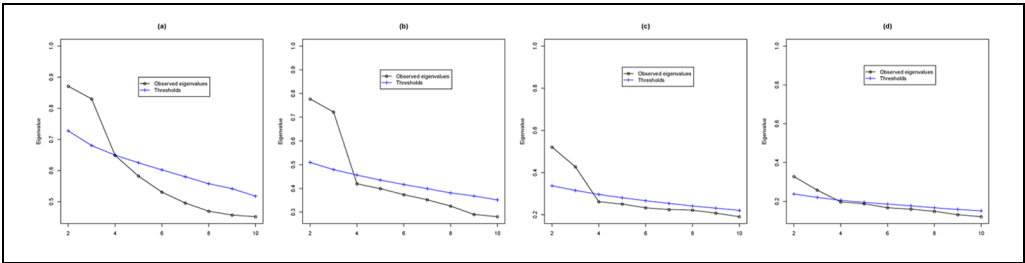


Figure 7. Observed eigenvalues versus the thresholds computed based on Monte Carlo samples from the reference distribution ($n = 100$).
Note. (a) $K = 5$; (b) $K = 10$; (c) $K = 20$; (d) $K = 30$.

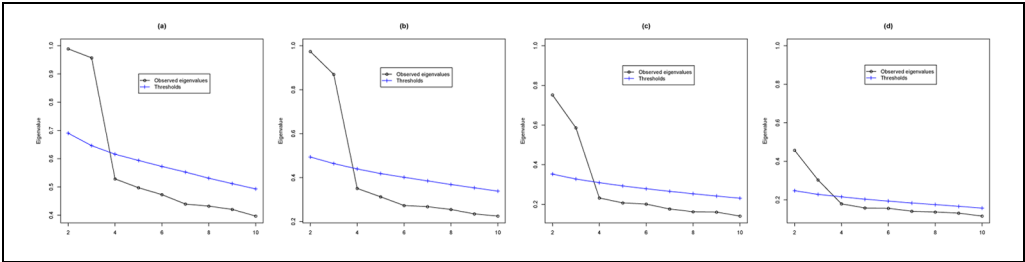


Figure 8. Observed eigenvalues versus the thresholds computed based on Monte Carlo samples from the reference distribution ($n = 400$).
Note. (a) $K = 5$; (b) $K = 10$; (c) $K = 20$; (d) $K = 30$.

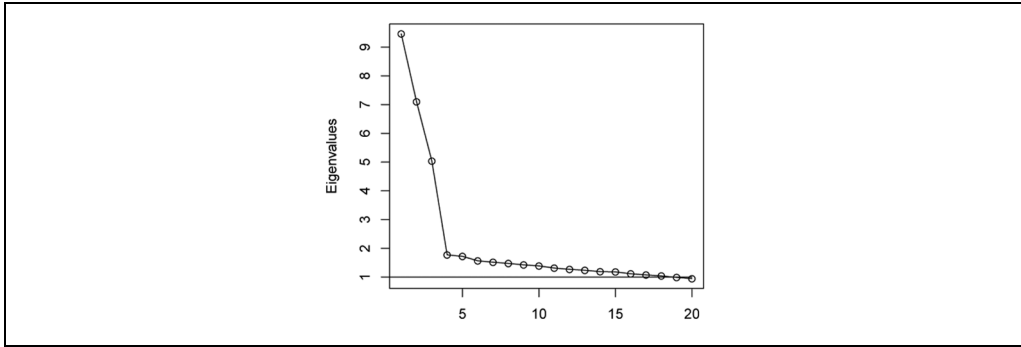


Figure 9. The top eigenvalues of the estimated polychoric correlation matrix for a simulated data set from the latent factor model, with sample size $n = 400$.

Under this measure, two reversely worded items that measure the same construct will be regarded as dissimilar and thus not clustered together. However, the proposed method is invariant due to the use of the absolute value of Kendall's tau statistic as a similarity measure.

Comparing With Exploratory Factor Analysis

When data are generated from the latent factor model as above, exploratory factor analysis tends to work in terms of choosing both the number of factors (clusters) and classification. Exploratory factor analysis is applied to the data set with $N = 400$ in the section “Illustration,” using the R package “psych” (Revelle, 2016). The top eigenvalues of the estimated polychoric correlation are shown in Figure 9, and the parallel analysis of exploratory factor analysis (Horn, 1965) selects three factors. Under the “oblimin” rotation, a simple structure consistent with the generating model is observed, where “oblimin” is a widely used oblique rotational method (Costello & Osborne, 2009). Clustering the estimated loadings using the K-means algorithm only misclassifies one item.

The above result is not surprising, as the data are generated from the model assumed by exploratory factor analysis. However, what if item clusters are generated from a different model? To answer this, the authors simulate an example from an unfolding model, which is widely used in attitude measurement (Roberts, Donoghue, & Laughlin, 2000). The data are generated from a multidimensional unfolding model (DeSarbo & Hoffman, 1986), with the following specifications: the number of clusters $M = 5$, the number of items $J = 60$, sample size $n = 1,000$, and $P(\pi_j = m) = 1/M$, where π_1, \dots, π_J are the cluster membership. Furthermore, each item is presented by an ideal point, denoted by $\delta_j = (\delta_{j1}, \delta_{j2})^\top$, on a two-dimensional space, such that

$$\delta_j | \pi_j = m \sim N\left(\mu_m, \begin{pmatrix} 0.25^2 & 0 \\ 0 & 0.25^2 \end{pmatrix}\right),$$

where $\mu_1 = (1, 0)^\top$, $\mu_2 = (0, 1)^\top$, $\mu_3 = (-1, 0)^\top$, $\mu_4 = (0, -1)^\top$, and $\mu_5 = (0, 0)^\top$. The generated ideal points of the items are shown in Figure 10, where clusters are indicated by colors. Each individual is represented by an ideal point, denoted by θ_i , on the same space. It is assumed that θ_i follows a standard multivariate normal. The responses are generated by

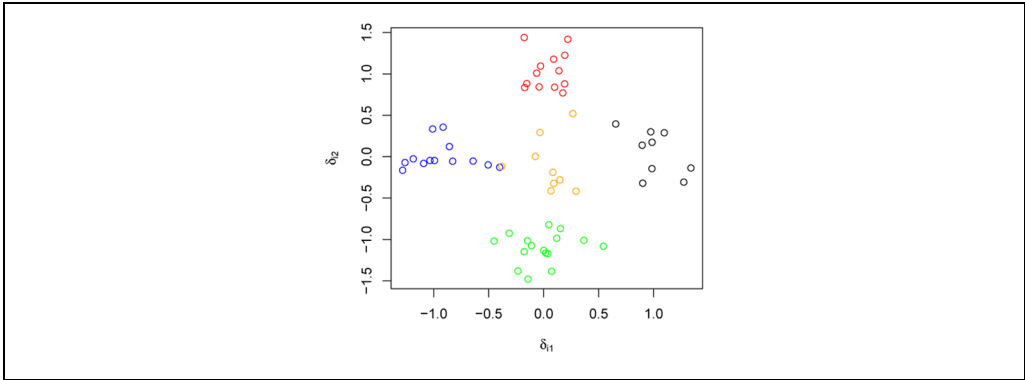


Figure 10. A scatter plot of the ideal points of items, where item clusters are marked by different colors.

$$P(R_{ij} = 1 | \delta_j, \theta_i) = \frac{1}{1 + e^{\|\theta_i - \delta_j\|^2}},$$

where $\|\theta_i - \delta_j\|$ is the Euclidian distance between individual i and item j in the ideal point space. The closer the two ideal points, the more likely the individual agrees with the item ($R_{ij} = 1$).

The spectral clustering method correctly selects the number of clusters, and only three points are misclassified. The authors then analyze the data with exploratory factor analysis. As the item clusters are generated from a different model, the estimated number of factors is no longer a good indication of the number of clusters and the extracted factors are difficult to interpret. The top eigenvalues of the estimated polychoric correlation matrix are shown in Figure 11, and parallel analysis selects four latent factors. Although the exploratory factor analysis model is misspecified, the estimated factor loadings still contain information about the clusters. That is, if we fit five-dimensional and four-dimensional exploratory factor models and then run a K-means algorithm to the estimated factor loadings under the “oblimin” rotation, the numbers of misclassified items are 5 and 6, respectively.

Finally, the authors replicate the comparison of the above results by simulating 100 data sets for each setting, and the corresponding results are shown in Table 1. For spectral clustering, they check the estimated number of clusters and the classification error when the number of clusters is given. For exploratory factor analysis, the estimated number of factors from parallel analysis is used as the estimated number of clusters, and the classification is from K-means clustering of the estimated loadings of an M -dimensional factor model. For the estimation of the number of clusters, they report the percentage of correct estimation over 100 replications, and for the classification, they report the averaged classification error rate and its standard error. From Table 1, it is evident that both methods perform well when the true model is a latent factor model. When data are generated from an unfolding model, the parallel analysis of exploratory factor analysis is not intended to estimate the number of clusters and underestimates the number of clusters all the time, while the spectral clustering method correctly identifies the number of clusters 76% of the time. The misclassification error rates are similar for both methods, with the spectral clustering method performing slightly better.

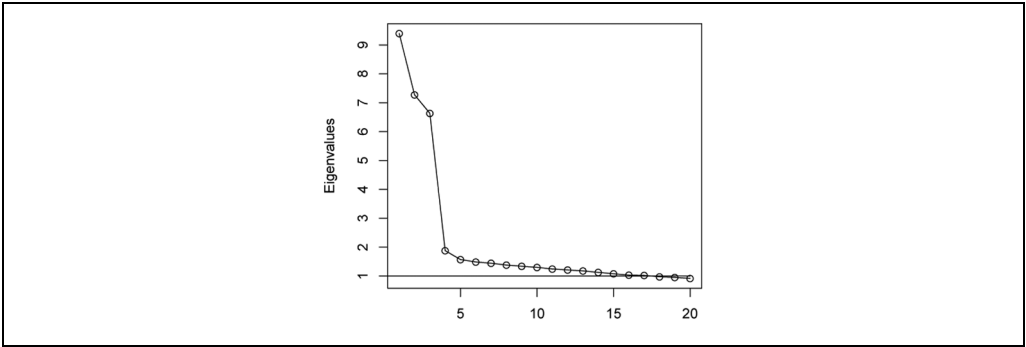


Figure 11. The top eigenvalues of the estimated polychoric correlation matrix for a simulated data set from an unfolding model.

Table 1. Results of Comparing Spectral Clustering and Exploratory Item Factor Analysis.

| | Spectral | | EFA | |
|-----------------|-----------|----------------|-----------|--------------------------|
| | # Cluster | Classification | # Cluster | Classification (oblimin) |
| Factor model | 100 | 1.6 (0.1) | 97 | 2.2 (0.3) |
| Unfolding model | 76 | 6.0 (0.4) | 0 | 7.1 (0.5) |

Note. “Spectral” and “EFA” represent spectral clustering and exploratory factor analysis, and “# cluster” and “classification” represent the estimation accuracy of the number of clusters and the classification error rate, respectively.

Application to EPQ-R Data

In this section, the authors apply the spectral clustering method to the EPQ-R (S. B. Eysenck et al., 1985). The study is initially a confirmatory study associated with three scales: Psychoticism/Socialization (P), Extraversion/Introversion (E), and Neuroticism/Stability (N). Briefly speaking, the P scale measures a personality pattern typified by aggressiveness and interpersonal hostility; the E scale measures an attitude type characterized by concentration of interest on the external object, and the N scale measures a personality type characterized by anxiety, fear, moodiness, worry, envy, frustration, jealousy, and loneliness. The data set contains a total of 79 items with 824 respondents. There are 32, 23, and 24 items on the P, E, and N scales, respectively. The specific questions and the score keys can be found in Appendix 1 and Table 4 of S. B. Eysenck et al. (1985). To facilitate the presentation, the items are reordered by scales. The new item IDs in this analysis and the corresponding ones in S. B. Eysenck et al. (1985) are presented in Table 2. In particular, Items 14 to 32 and 53 to 55 (under the current set of IDs) are reversely worded. The data have been analyzed using item response theory models, such as in Maydeu-Olivares and Liu (2015) and Y. Liu, Magnus, Quinn, and Thissen (in press). To the best knowledge of the authors, this is the first cluster analysis of EPQ-R items.

The three scales are first validated by cluster analysis. In Table 3, the results of spectral clustering are presented, assuming that there are three clusters, and the connectivity parameter *K* is chosen as 5, as suggested in Algorithm 3. The results turn out to be insensitive to the choice of *K*. The item clustering results are identical to the confirmatory structure mentioned in the

Table 2. New Item IDs Used in This Analysis and the Corresponding Ones in Appendix I of S. B. Eysenck, Eysenck, and Barrett (1985).

| | | | | | | | | | | | | | | | | |
|--------------------------------|----|----|----|----|----|----|----|-----|----|----|----|----|----|----|----|----|
| Psychoticism/Socialization (P) | | | | | | | | | | | | | | | | |
| New ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Original ID | 25 | 29 | 30 | 34 | 37 | 42 | 48 | 50 | 56 | 73 | 75 | 91 | 95 | 2 | 5 | 7 |
| New ID | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
| Original ID | 9 | 12 | 14 | 18 | 21 | 41 | 54 | 59 | 64 | 68 | 79 | 81 | 85 | 88 | 96 | 99 |
| Extraversion/Introversion (E) | | | | | | | | | | | | | | | | |
| New ID | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 |
| Original ID | 1 | 6 | 11 | 16 | 20 | 28 | 36 | 40 | 45 | 51 | 55 | 58 | 61 | 63 | 67 | 69 |
| New ID | 49 | 50 | 51 | 52 | 53 | 54 | 55 | | | | | | | | | |
| Original ID | 72 | 78 | 90 | 94 | 24 | 33 | 47 | | | | | | | | | |
| Neuroticism/Stability (N) | | | | | | | | | | | | | | | | |
| New ID | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 |
| Original ID | 3 | 8 | 13 | 17 | 22 | 26 | 31 | 35 | 38 | 43 | 46 | 52 | 60 | 65 | 70 | 74 |
| New ID | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | | | | | | | | |
| Original ID | 76 | 80 | 83 | 84 | 87 | 92 | 97 | 100 | | | | | | | | |

Table 3. Spectral Clustering With Three Clusters.

| | Item |
|---|---|
| 1 | 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 30, 31, 32 |
| 2 | 14 , 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 |
| 3 | 4, 15, 16, 24, 29 , 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79 |

Note. The boldfaced values are items that are misclassified.

previous paragraph, except for six items: Item 14 “Do you stop to think things over before doing anything?” from the P scale is classified as Cluster 2 that corresponds to the E scale, and Items 4 “Do you have enemies who want to harm you?” 15 “Do you take much notice of what people think?” 16 “Would being in debt worry you?” 24 “Does it worry you if you know there are mistakes in your work?” and 29 “Can you on the whole trust people to tell the truth?” from the P scale are classified as Cluster 3 that corresponds to the N scale. Figure 12 presents the graph features corresponding to the second and the third eigenvectors of the Laplacian matrix, where the items from the P, E, and N scales are marked in black, red, and blue, respectively. Items 14, 4, 15, 16, 24, and 29 are on the boundary, suggesting that Item 14 is similar to both items from the P and E scales, and Items 4, 15, 16, 24, and 29 are similar to both items from the P and N scales. The algorithm reconstructs the three scales reasonably well without prior information of the confirmatory structure, given that $M = 3$ clusters are enforced.

The authors compare the results above with results obtained from the K-means method. Simply applying the K-means method completely destroys the three-scale structure of EPQ-R, as shown in Table 4, mainly due to the existence of reversely worded items. In particular, it is observed that each cluster is a mixture of items from the three scales. If the data are preprocessed by manually flipping the responses to the reversely worded items, the K-means method produces reasonable results. However, even with that, the result of K-means method is still not

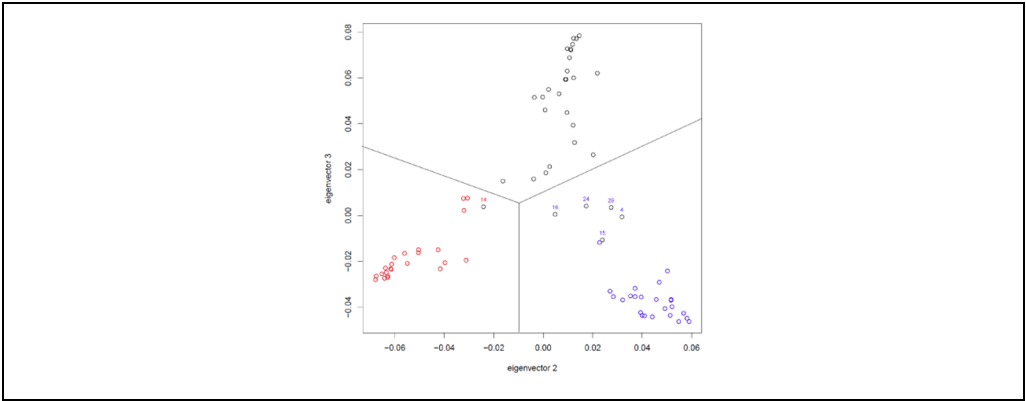


Figure 12. Second and third eigenvectors.

Table 4. K-Means Clustering With Three Clusters.

| Item | |
|------|--|
| 1 | 1, 3, 4, 5, 7, 9, 10, 11, 12, 13, 53, 54, 55, 59, 63, 66, 68, 70, 72, 74, 75 |
| 2 | 2, 6, 8, 30, 33, 34, 40, 41, 42, 45, 46, 47, 48, 49, 50, 52, 56, 57, 61, 62, 64, 65, 67, 69, 71, 73, 79 |
| 3 | 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 35, 36, 37, 38, 39, 43, 44, 51, 58, 60, 76, 77, 78 |

Table 5. K-Means Clustering With Three Clusters When the Responses to the Reversely Worded Items Are Flipped.

| Item | |
|------|---|
| 1 | 1, 3, 4, 5, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 42, 68 |
| 2 | 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 47, 48, 49, 50, 51, 52, 53, 54, 55, 58, 77, 78 |
| 3 | 2, 6, 8, 20, 30, 45, 46 , 56, 57, 59, 60, 61, 62, 63, 64, 65, 66, 67, 69, 70, 71, 72, 73, 74, 75, 76, 79 |

Note. The boldfaced values are items that are misclassified.

as good as that of spectral clustering. In particular, as shown in Table 5, the three clusters correspond to the P, E, and N scales, respectively, but there are 12 items misclassified. These items include 42, 68, 58, 77, 78, 2, 6, 8, 20, 30, 45, and 46. The authors point out that flipping the responses to a subset of items does not change the result of the proposed method, because the absolute values of Kendall’s tau statistics do not change. The better performance of spectral clustering is possibly due to that the similarity measure based on Kendall’s tau statistic captures the underlying structure of the items better than that based on the Euclidian distance. In addition, it is often unknown which items are reversely worded, especially at the early stage of inventory construction. In such a situation, the proposed method still works, while the K-means is likely to fail.

The authors extend the analysis to the situation when the number of clusters is unknown. They choose $K=5$ as suggested in Algorithm 3. The result of parallel analysis is shown in Figure 13, where the suggested number of clusters is 8. The clustering results for $M=8$ and

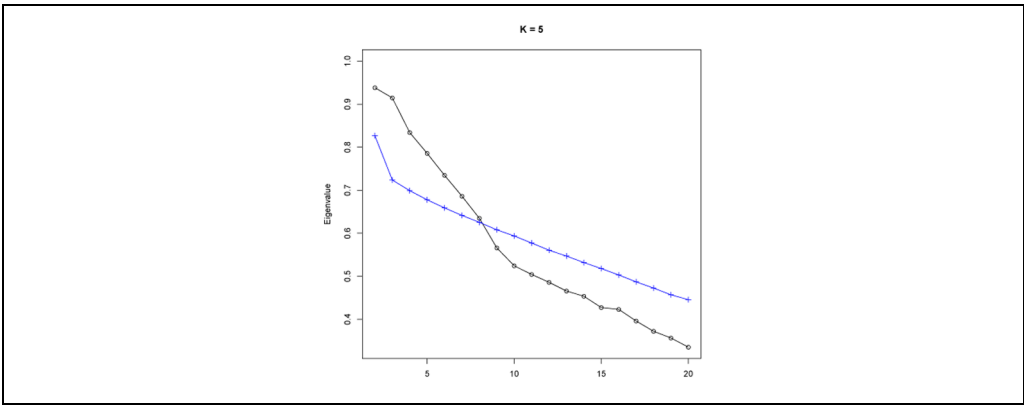


Figure 13. Result from parallel analysis for the EPQ-R data.
Note. EPQ-R = Eysenck Personality Questionnaire–Revised.

Table 6. Spectral Clustering With Eight Clusters; Connectivity Parameter $K = 5$.

| | Item | Interpretation |
|---|--|----------------------------------|
| 1 | 1, 2, 6, 7, 8, 11, 19, 20, 21, 22, 25, 30, 31 | Irresponsible |
| 2 | 3, 4, 9, 10, 12, 13, 18, 23, 26, 27, 29, 32 | Hostile and unsympathetic |
| 3 | 5, 14, 15, 16, 17, 24, 28, 40, 45, 48 | Egocentric and impulsive |
| 4 | 33, 34, 35, 41, 42, 46, 47, 49, 50, 52, 53, 55 | Expressive, active, and dominant |
| 5 | 36, 37, 38, 39, 43, 44, 51, 54 | Sociable |
| 6 | 56, 57, 59, 61, 69, 70, 71, 75, 77, 79 | Moody and depressed |
| 7 | 58, 60, 73, 76, 78 | Anxious and tense |
| 8 | 62, 63, 64, 65, 66, 67, 68, 72, 74 | Obsessive |

Note. The boldfaced values are three E scale items placed in a cluster with the rest of the items from the P scale.

$K = 5$ are shown in Table 6. The eight clusters shown in Table 6 are basically finer scales of the original three scales in the confirmatory analysis. In particular, the P scale is split into three subscales, the E scale is split into two subscales, and the N scale is split into three subscales, except that Items 40, 45, and 48 from the E scale are classified into Cluster 3, a subscale of Psychoticism.

The authors believe that the results are very reasonable, as the items in each cluster summarize different aspects of a scale (see Table 6). For example, items in Cluster 1 are about irresponsibility, and those in Cluster 2 are about hostility and unsympathy, and so on. In addition, it is reasonable to see Items 40 “Would you call yourself happy-go-lucky?” 45 “Have people said that you sometimes act too rashly?” and 48 “Do you often make decisions on the spur of the moment?” in Cluster 3, because they are all about impulsiveness, and are very similar to Items 14 “Do you stop to think things over before doing anything?” and 28 “Do you generally ‘look before you leap?’” that are in the same cluster.

Summary and Discussion

The authors develop a spectral algorithm for the cluster analysis of large-scale item response data. A key component of the algorithm is the construction of the similarity matrix based on

Kendall's tau statistic. This construction makes the algorithm applicable to both binary and ordinal response data and effective when data have a significant amount of missing or incomplete responses. The proposed algorithm is computationally efficient and easy to implement. The authors also propose a method to select the number of clusters. The method is easy to implement and performs well according to the simulation studies and real data application. In the application to EPQ-R data, eight clusters are suggested, and the corresponding clusters receive good interpretation. The authors point out that the proposed spectral clustering algorithm is an exploratory tool. To be used as assessment scales, the resulting item clusters need to be confirmed or improved upon by domain experts and by standard analysis based on latent factor models.

The proposed method is applicable to analyzing item response and other behavioral data from many areas, including personality, psychiatry, marketing, political science, and so on. Thanks to its computational efficiency, flexibility to different data structures, and reliable performance, the proposed method could serve as a useful exploratory tool for analyzing data that are heterogeneous and of large scale.

There are remaining issues that need to be addressed: First, the Kendall's tau similarity measure can be replaced by a different measure of similarity, such as the polychoric correlation. Second, the theoretical properties of the proposed method need to be established. In particular, if data are generated from a latent factor model with a simple structure, it is not difficult to show that the proposed method consistently clusters items that measure the same latent factor, as the sample size goes to infinity. A more challenging asymptotic setting is that the sample size, the number of items, and the number of factors all grow to infinity. It is of interest to investigate when the proposed algorithm works and when it does not. Finally, it is likely that there are items that measure multiple latent factors. For this, it is necessary to develop spectral clustering algorithms that allow observations to belong to multiple clusters.

Acknowledgment

We would like to thank Dr. Paul Barrett for letting us use the Eysenck Personality Questionnaire-Revised (EPQ-R) data.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by NSF Grant SES-1323977, NSF Grant IIS-1633360, Army Research Office Grant W911NF-15-1-0159, Institute of Education Sciences Grant R305D160010, NSA Grant H98230-16-1-0299, and NIH Grant R01GM047845.

References

- Azran, A., & Ghahramani, Z. (2006). Spectral methods for automatic multiscale data clustering. In A. Fitzgibbon, C. J. Taylor & Y. LeCun (Eds.), *2006 IEEE computer society conference on computer vision and pattern recognition* (Vol. 1. pp. 190-197). Los Alamitos, CA: IEEE Computer Society Press.
- Blashfield, R. K. (1980). The growth of cluster analysis: Tryon, Ward, and Johnson. *Multivariate Behavioral Research*, 15, 439-458.

- Blashfield, R. K., & Aldenderfer, M. S. (1988). The methods and problems of cluster analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 447-473). New York, NY: Plenum Press.
- Borgen, F. H., & Barnett, D. C. (1987). Applying cluster analysis in counseling psychology research. *Journal of Counseling Psychology*, 34, 456-468.
- Brito, M., Chavez, E., Quiroz, A., & Yukich, J. (1997). Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection. *Statistics & Probability Letters*, 35, 33-42.
- Cattell, R. B., & Kline, P. E. (1977). *The scientific analysis of personality and motivation*. New York, NY: Penguin Books.
- Chen, Y., Li, X., Liu, J., & Ying, Z. (2016). *A fused latent and graphical model for multivariate binary data*. ArXiv e-prints. Retrieved from <https://arxiv.org/pdf/1606.08925.pdf>
- Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, 110, 850-866.
- Chi, J. T., Chi, E. C., & Baraniuk, R. G. (2016). K-POD: A method for k-means clustering of missing data. *The American Statistician*, 70, 91-99.
- Chiu, C.-Y., & Douglas, J. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification*, 30, 225-250.
- Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74, 633-665.
- Chung, F. R. (1997). *Spectral graph theory*. Providence, RI: American Mathematical Society.
- Costello, A. B., & Osborne, J. W. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7), 1-9.
- DeSarbo, W. S., & Hoffman, D. L. (1986). Simple and weighted unfolding threshold models for the spatial representation of binary choice data. *Applied Psychological Measurement*, 10, 247-264.
- Dhillon, I. S. (2001). Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 269-274). New York, NY: Association for Computing Machinery.
- Eysenck, H. J. (1982). *Personality, genetics and behavior: Selected papers*. New York, NY: Praeger.
- Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences*, 6, 21-29.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96.
- Hartigan, J. A. (1975). *Clustering algorithms*. New York, NY: John Wiley.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Howell, D. C. (2012). *Statistical methods for psychology*. Belmont, CA: Cengage Learning.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31, 264-323.
- Kendall, M. G., & Gibbons, J. D. (1990). *Rank correlation methods* (5th ed.). London, England: Edward Arnold.
- Kolaczyk, E. D. (2009). *Statistical analysis of network data: Methods and models*. New York, NY: Springer.
- Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology*, 48, 339-358.
- Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement*, 36, 548-564.
- Liu, J., Xu, G., & Ying, Z. (2013). Theory of the self-learning Q-matrix. *Bernoulli*, 19, 1790-1817.
- Liu, Y., Magnus, B., Quinn, H., & Thissen, D. (in press). Multidimensional item response theory. In I. P. D. Hughes & T. Booth (Eds.), *Handbook of psychometric testing*. Hoboken, NJ: Wiley-Blackwell.

- Loevinger, J., Gleser, G. C., & DuBois, P. H. (1953). Maximizing the discriminating power of a multiple-score test. *Psychometrika*, 18, 309-317.
- Maydeu-Olivares, A., & Liu, Y. (2015). Item diagnostics in multivariate discrete data. *Psychological Methods*, 20, 276-292.
- Ng, A. Y., Jordan, M. I., Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In T. Dietterich, S. Becker, Z. Ghahramani (Eds.), *Advances in neural information processing systems 14* (Vol. 2., pp. 849-856). Cambridge, MA: MIT Press.
- Park, Y., & Lee, Y. (2011). Diagnostic cluster analysis of mathematics skills. *IERI Monograph Series: Issues and Methodologies in Large-scale Assessments*, 4, 75-107.
- Revelle, W. (2016). *An overview of the psych package*. Retrieved from <http://personality-project.org/r/overview.pdf>
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, 24, 3-32.
- Rohe, K., Chatterjee, S., & Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39, 1878-1915.
- Roussos, L. A., Stout, W. F., & Marden, J. I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement*, 35, 1-30.
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 888-905.
- Sprent, P., & Smeeton, N. C. (2016). *Applied nonparametric statistical methods*. Boca Raton, FL: CRC Press.
- Stewart, G. W., & Sun, J.-G. (1990). *Matrix perturbation theory*. New York, NY: Academic Press.
- Sun, J., Chen, Y., Liu, J., Ying, Z., & Xin, T. (2016). Latent variable selection for multidimensional item response theory models via L_1 regularization. *Psychometrika*, 84, 921-939.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 63, 411-423.
- von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17, 395-416.
- Willse, J., Henson, R., & Templin, J. (2007, April). *Using sum scores or IRT in place of cognitive diagnosis models: Can existing or more familiar models do the job?* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Xing, J., Watkins, W. S., Shlien, A., Walker, E., Huff, C. D., & Witherspoon, D. J., . . . Jorde, L. B. (2010). Toward a more uniform sampling of human genetic diversity: A survey of worldwide populations by high-density genotyping. *Genomics*, 96, 199-210.
- Zhang, J. (2007). Conditional covariance theory and detect for polytomous items. *Psychometrika*, 72, 69-91.
- Zhang, J. (2013). A procedure for dimensionality analyses of response data from various test designs. *Psychometrika*, 78, 37-58.
- Zhang, J., & Stout, W. (1999). The theoretical detect index of dimensionality and its application to approximate simple structure. *Psychometrika*, 64, 213-249.